1. PROBABILITY

Revised September 2009 by G. Cowan (RHUL).

The following is a much-shortened version of Sec. 31 of the full *Review*. Equation, section, and figure numbers follow the *Review*.

1.2. Random variables

• *Probability density function* (p.d.f.): x is a *random variable*.

Continuous: $f(x;\theta)dx$ = probability x is between x to $x+dx$, given parameter(s) θ;

Discrete: $f(x;\theta)$ = probability of x given θ.

• *Cumulative distribution function*:

$$F(a) = \int_{-\infty}^{a} f(x)\ dx\ . \tag{1.6}$$

Here and below, if x is discrete-valued, the integral is replaced by a sum. The endpoint a is induced in the integral or sum.

• *Expectation values*: Given a function u:

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\ f(x)\ dx\ . \tag{1.7}$$

• *Moments*:

$$n\text{th moment of a random variable: } \alpha_n = E[x^n]\ , \tag{1.8a}$$

$$n\text{th central moment: } m_n = E[(x-\alpha_1)^n]\ . \tag{1.8b}$$

$$\text{Mean:} \qquad \mu \equiv \alpha_1\ . \tag{1.9a}$$

$$\text{Variance:} \qquad \sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2\ . \tag{1.9b}$$

Coefficient of skewness: $\gamma_1 \equiv m_3/\sigma^3$.
Kurtosis: $\gamma_2 = m_4/\sigma^4 - 3$.
Median: $F(x_{\rm med}) = 1/2$.

• *Marginal* p.d.f.: Let x,y be two random variables with joint p.d.f. $f(x,y)$.

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\ dy\ ; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x,y)\ dx\ . \tag{1.10}$$

• *Conditional* p.d.f.:

$$f_4(x|y) = f(x,y)/f_2(y)\ ; \quad f_3(y|x) = f(x,y)/f_1(x)\ .$$

• *Bayes' theorem*:

$$f_4(x|y) = \frac{f_3(y|x)\,f_1(x)}{f_2(y)} = \frac{f_3(y|x)\,f_1(x)}{\int f_3(y|x')\,f_1(x')\ dx'}\ . \tag{1.11}$$

• *Correlation coefficient and covariance*:

$$\mu_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) \; dx \; dy \; , \tag{1.12}$$

$$\rho_{xy} = E\left[(x-\mu_x)\left(y-\mu_y\right)\right] / \sigma_x \, \sigma_y \equiv \mathrm{cov}\,[x,y] / \sigma_x \, \sigma_y \; ,$$

$$\sigma_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_x)^2 \; f(x,y) \; dx \; dy \; . \text{ Note } \rho_{xy}^2 \leq 1.$$

• *Independence*: x,y are independent if and only if $f(x,y) = f_1(x) \cdot f_2(y)$; then $\rho_{xy} = 0$, $E[u(x) \; v(y)] = E[u(x)] \; E[v(y)]$ and $V[x+y] = V[x] + V[y]$.

• *Change of variables*: From $\boldsymbol{x} = (x_1, \ldots, x_n)$ to $\boldsymbol{y} = (y_1, \ldots, y_n)$: $g(\boldsymbol{y}) = f\left(\boldsymbol{x}(\boldsymbol{y})\right) \cdot |J|$ where $|J|$ is the absolute value of the determinant of the Jacobian $J_{ij} = \partial x_i / \partial y_j$. For discrete variables, use $|J| = 1$.

1.3. Characteristic functions

Given a pdf $f(x)$ for a continuous random variable x, the characteristic function $\phi(u)$ is given by (31.6). Its derivatives are related to the algebraic moments of x by (31.7).

$$\phi(u) = E\left[e^{iux}\right] = \int_{-\infty}^{\infty} e^{iux} f(x) \; dx \; . \tag{1.17}$$

$$i^{-n} \; \frac{d^n \phi}{du^n}\bigg|_{u=0} = \int_{-\infty}^{\infty} x^n f(x) \; dx = \alpha_n \; . \tag{1.18}$$

If the p.d.f.s $f_1(x)$ and $f_2(y)$ for independent random variables x and y have characteristic functions $\phi_1(u)$ and $\phi_2(u)$, then the characteristic function of the weighted sum $ax + by$ is $\phi_1(au)\phi_2(bu)$. The additional rules for several important distributions (*e.g.*, that the sum of two Gaussian distributed variables also follows a Gaussian distribution) easily follow from this observation.

1.4. Some probability distributions

See Table 1.1.

1.4.2. *Poisson distribution*:

The Poisson distribution $f(n;\nu)$ gives the probability of finding exactly n events in a given interval of x (*e.g.*, space or time) when the events occur independently of one another and of x at an average rate of ν per the given interval. The variance σ^2 equals ν. It is the limiting case $p \to 0$, $N \to \infty$, $Np = \nu$ of the binomial distribution. The Poisson distribution approaches the Gaussian distribution for large ν.

For example, a large number of radioactive nuclei of a given type will result in a certain number of decays in a fixed time interval. If this interval is small compared to the mean lifetime, then the probability for a given nucleus to decay is small, and thus the number of decays in the time interval is well modeled as a Poisson variable.

1.4.3. ***Normal or Gaussian distribution***:

Its cumulative distribution, for mean 0 and variance 1, is usually tabulated as the *error function*

$$F\,(x;0,1)=\tfrac{1}{2}\left[1+\mathrm{erf}\left(x/\sqrt{2}\right)\right]\ . \tag{1.24}$$

For mean μ and variance σ^2, replace x by $(x-\mu)/\sigma$. The error function is acessible in libraries of computer routines such as CERNLIB.

$P(x \text{ in range } \mu\pm\sigma)=0.6827$,

$P(x \text{ in range } \mu\pm 0.6745\sigma)=0.5$,

$E[|x-\mu|]=\sqrt{2/\pi}\sigma=0.7979\sigma$,

half-width at half maximum $=\sqrt{2\ln 2}\cdot\sigma=1.177\sigma$.

Table 1.1. Some common probability density functions, with corresponding characteristic functions and means and variances. In the Table, $\Gamma(k)$ is the gamma function, equal to $(k-1)!$ when k is an integer.

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu}-e^{iau}}{(b-a)iu}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;N,p) = \dfrac{N!}{r!(N-r)!}\; p^r q^{N-r}$ $r = 0,1,2,\ldots,N\,; \quad 0 \le p \le 1\,; \quad q = 1-p$	$(q+pe^{iu})^N$	Np	Npq
Poisson	$f(n;\nu) = \dfrac{\nu^n e^{-\nu}}{n!}\;; \quad n = 0,1,2,\ldots\,; \quad \nu > 0$	$\exp[\nu(e^{iu}-1)]$	ν	ν
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\;\exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\,; \quad -\infty < \mu < \infty\,; \quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},V) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert V\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty; \quad -\infty < \mu_j < \infty; \quad \lvert V\rvert > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V \boldsymbol{u}\right]$	$\boldsymbol{\mu}$	V_{jk}
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\;; \quad z \ge 0$	$(1-2iu)^{-n/2}$	n	$2n$
Student's t	$f(t;n) = \dfrac{1}{\sqrt{n\pi}}\;\dfrac{\Gamma[(n+1)/2]}{\Gamma(n/2)}\left(1+\dfrac{t^2}{n}\right)^{-(n+1)/2}$ $-\infty < t < \infty\,; \quad n$ not required to be integer	—	0 for $n > 1$	$n/(n-2)$ for $n > 2$
Gamma	$f(x;\lambda,k) = \dfrac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}\;; \quad 0 \le x < \infty\,;$ k not required to be integer	$(1-iu/\lambda)^{-k}$	k/λ	k/λ^2

For n Gaussian random variables $\boldsymbol{x}_i$, the joint p.d.f. is the multivariate Gaussian:

$$f(\boldsymbol{x};\boldsymbol{\mu},V) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],\ |V|>0\,. \tag{1.27}$$

V is the $n\times n$ *covariance matrix*; $V_{ij} \equiv E[(x_i-\mu_i)(x_j-\mu_j)] \equiv \rho_{ij}\,\sigma_i\,\sigma_j$, and $V_{ii} = V[x_i]$; $|V|$ is the determinant of V. For $n=2$, $f(\boldsymbol{x};\boldsymbol{\mu},V)$ is

$$f(x_1,x_2;\ \mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \ \exp\left\{\frac{-1}{2(1-\rho^2)}\right.$$

$$\left.\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}\,. \tag{1.28}$$

The marginal distribution of any x_i is a Gaussian with mean μ_i and variance V_{ii}. V is $n\times n$, symmetric, and positive definite. Therefore for any vector $\boldsymbol{X}$, the quadratic form $\boldsymbol{X}^T V^{-1}\boldsymbol{X} = C$, where C is any positive number, traces an n-dimensional ellipsoid as $\boldsymbol{X}$ varies. If $X_i = x_i - \mu_i$, then C is a random variable obeying the χ^2 distribution with n degrees of freedom, discussed in the following section. The probability that $\boldsymbol{X}$ corresponding to a set of Gaussian random variables x_i lies outside the ellipsoid characterized by a given value of C ($=\chi^2$) is given by $1 - F_{\chi^2}(C;n)$, where F_{χ^2} is the cumulative χ^2 distribution. This may be read from Fig. 32.1. For example, the "s-standard-deviation ellipsoid" occurs at $C = s^2$. For the two-variable case ($n=2$), the point $\boldsymbol{X}$ lies outside the one-standard-deviation ellipsoid with 61% probability. The use of these ellipsoids as indicators of probable error is described in Sec. 32.3.2.4; the validity of those indicators assumes that $\boldsymbol{\mu}$ and V are correct.

1.4.4. χ^2 *distribution*:

If $x_1,\ldots,x_n$ are independent Gaussian random variables, the sum $z = \sum_{i=1}^n (x_i-\mu_i)^2/\sigma_i^2$ follows the χ^2 p.d.f. with n *degrees of freedom*, which we denote by $\chi^2(n)$. More generally, for n correlated Gaussian variables as components of a vector $\boldsymbol{X}$ with covariance matrix V, $z = \boldsymbol{X}^T V^{-1}\boldsymbol{X}$ follows $\chi^2(n)$ as in the previous section. For a set of z_i, each of which follows $\chi^2(n_i)$, $\sum z_i$ follows $\chi^2(\sum n_i)$. For large n, the χ^2 p.d.f. approaches a Gaussian with mean $\mu = n$ and variance $\sigma^2 = 2n$.

The χ^2 p.d.f. is often used in evaluating the level of compatibility between observed data and a hypothesis for the p.d.f. that the data might follow. This is discussed further in Sec. 32.2.2 on tests of goodness-of-fit.

1.4.6. *Gamma distribution*:

For a process that generates events as a function of x (*e.g.*, space or time) according to a Poisson distribution, the distance in x from an arbitrary starting point (which may be some particular event) to the k^{th} event follows a *gamma* distribution, $f(x;\lambda,k)$. The Poisson parameter μ is λ per unit x. The special case $k=1$ (*i.e.*, $f(x;\lambda,1)=\lambda e^{-\lambda x}$) is called the *exponential* distribution. A sum of k' exponential random variables x_i is distributed as $f(\sum x_i;\lambda,k')$.

The parameter k is not required to be an integer. For $\lambda=1/2$ and $k=n/2$, the gamma distribution reduces to the $\chi^2(n)$ distribution.

See the full *Review* for further discussion and all references.